REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
14-11-2007

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*
1/5/2005 - 9/30/2007

4. TITLE AND SUBTITLE
Experimental Evaluation of Performance Feedback Using the Dismounted Infantry Virtual After Action Review System

Long Range Navy and Marine Corps Science & Technology Program

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-05-1-0171

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
250841

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
James P. Bliss, Ph.D.
Ernesto A. Bustamante
Corey K. Fallon
Katherine M. Ely

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Old Dominion University Research Foundation
P. O. Box 6369
Norfolk, VA 23508

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Training soldiers for modern warfare is a critical and complex undertaking. An important tool used to train soldiers is the after action review (AAR), during which soldier trainees review positive and negative aspects of past training performances. The Dismounted Infantry Virtual After Action Review System (DIVAARS), created by the Army Research Institute and the Institute for Simulation and Training, allows soldiers to review past immersive training exercises. Our experimental team planned to use DIVAARS to study variables that may influence the effectiveness of performance feedback in an AAR. Planned experiments were to isolate and evaluate the roles of feedback timing, content, and teamed review on AAR quality and comprehensiveness and reviewers' situation awareness and perceived workload. Unfortunately, due to complications surrounding the acquisition of active duty military participants, we were unable to complete our proposed experiments. The following report describes the results of our literature review and the tasks we were able to complete in support of ARI/PEOSTRI.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU		James P. Bliss, Ph.D.
					19b. TELEPHONE NUMBER *(Include area code)* 757-683-4051

FINAL REPORT

Experimental Evaluation of Performance Feedback Using the Dismounted Infantry Virtual After Action Review System

Long Range Navy and Marine Corps Science & Technology Program

James P. Bliss, Ph.D., Principal Investigator
Ernesto A. Bustamante
Corey K. Fallon
Katherine M. Ely

Old Dominion University

Abstract

Training soldiers for modern warfare is a critical and complex undertaking. Military tasks have in many ways transcended basic responsibilities of moving, communicating, and employing weapons. It is crucial that soldiers maintain effective situation awareness to understand the cascading effects of the actions of their own and other units and to comprehend how changes in resources might affect mission success. An important tool used to train soldiers is the after action review (AAR), during which soldier trainees review positive and negative aspects of past training performances. The Dismounted Infantry Virtual After Action Review System (DIVAARS), created by the Army Research Institute and the Institute for Simulation and Training, allows soldiers to review past immersive training exercises. In the research described here, the experimental team planned to use DIVAARS to study variables that may influence the effectiveness of performance feedback given within the context of an AAR. Several planned experiments conducted over a three-year period were to isolate and evaluate the roles of feedback timing, content, and teamed review on AAR quality and comprehensiveness. Data were also to be collected regarding reviewers' situation awareness and perceived workload. Ultimately, we expected to contribute to existing knowledge concerning performance feedback, and to improve subsequent versions of the DIVAARS program. Unfortunately, due to complications surrounding the acquisition of active duty military participants, we were unable to complete any proposed experiments. The following report describes the results of our literature review and the tasks we were able to complete in support of ARI/PEOSTRI.

Statement of Work

Statement of the Problem – For years, researchers and practitioners have devoted effort to devise effective methods to train military forces. The use of part- and whole-task simulation has contributed to this goal by allowing trainers to create representations of tasks that are excessively dangerous and complex for conventional training. Simulation of military tasks has evolved over the years from relatively crude representations of simple tasks for individual training, to complex representations of multi-user tasks. Yet, a number of core issues have remained central to the success of such training. These include task presentation methods, unbiased and meaningful evaluations of task performances, transfer of trained principles to performance in the field, and the fidelity of task representations.

In recent years, the military has increased its use of artificial immersive environments to represent complex tasks and situations during training. Virtual environment technology, though crude at first, has grown into a viable medium for task instruction and practice. Virtual environments offer the advantages of task replication, comprehensive performance measurement, task modification and augmentation, safety, and reduced long-term expenditure. These advantages have encouraged the use of virtual environments not just for individual task training, but also for collective training of equipment-based and dismounted infantry operations.

One of the most important factors influencing the success of training is the knowledge of performance and knowledge of results given to trainees. Feedback is viewed as a rich source of information that, if judiciously presented, may result in maximized knowledge acquisition and retention. The specific provision of feedback following military training exercises has typically taken the form of after action reviews (AARs). Such reviews are ideally an opportunity for trainers and trainees to discuss aspects of the training exercise that were successful and unsuccessful. Unfortunately, AARs have not always been conducted according to a standardized or rigorous procedure. Often information is presented in an informal fashion, without adherence to empirically driven guidelines for feedback effectiveness.

As armed forces expand their use of immersive simulation technology for complex mission training, it is important that researchers address the existing barriers to effective after action reviews. The proposed research represents part of an initial coordinated effort to accomplish this goal. Using an established tool for the playback and review of recorded immersive training scenarios (the Dismounted Infantry Virtual After Action Review System, hereafter DIVAARS), we propose to investigate the influence of factors such as information timing, content, playback features, and teamed reviewing on AAR quality and comprehensiveness. We also propose to investigate the levels of situation awareness and perceived workload experienced by individual and teamed reviewers as they use DIVAARS.

We anticipate that our research will clarify the specific influence of feedback variations in an applied context. Ultimately, as the DIVAARS system makes the transition from a prototype to an applied tool, the results demonstrated here will ensure that soldiers receive immersive training that is based on established and confirmed methods of knowledge acquisition and transfer.

Background and Significance

Factors Affecting the Success of Military Training

There are a host of psychological factors that may influence the success of any immersive training technology used for training military personnel. Some of these issues relate to the cognitive factors necessary for soldiers to effectively process unfamiliar information. Other issues relate to the mechanics of the training process, including the content or material to be learned, and the implementation of the training. Included below is a discussion of some of the more salient issues.

Soldier Cognition - As noted by Proctor and Van Zandt's (1994), human learning relies heavily on the structure, function, and efficiency of our information processing system. Cognitive scientists have for years attempted to model and understand this process.

To effectively design training programs and formulate training technologies, is critical to understand how humans process information. If designers understand this process, they may structure task elements in a way that matches human expectations. As noted by Rasmussen (1986), accurate models of human information processing may contribute to the design of critical systems pertaining to complex tasks as diverse as industrial control and emergency management. For the communication process between humans and technology to be optimal, information must be presented in a way that the human operator can understand. Information processing models are also relevant for military training because curricula designers can manipulate the progression of task difficulty to ensure rapid learning and lengthy retention. Specific to the AAR process, designers can use knowledge of human information processing to diagnose training deficiencies and determine when task mastery has occurred.

According to Proctor and Van Zandt (1996), the basis for most modern conceptualizations of human information processing is a three-stage model. The first stage is a perceptual stage, followed by a cognitive stage, and then the action stage. In the perceptual stage, humans receive sensory information from the outside world. They then manipulate and analyze that information in the cognitive stage. Finally, they select an appropriate response to the information in the action stage.

Each of these stages holds implications for training. The design of training technology is a critical issue for the perceptual stage, so that highly critical stimuli are readily perceived during training exercises. If an important element of the training situation does not attract the soldier's attention, or if the soldier does not recognize the role of a critical stimulus such as a weapons cache, piece of terrain or vehicle, then the success of the overall training program may be compromised.

In the cognitive stage, trainees relate aspects of the training environment to information they already know. Recognition and evaluation are therefore both important during this stage of information processing. Trainers must develop ways to present information so that it matches trainees' frames of reference. A significant challenge for training designers here is understanding the notion of "situation awareness," or the extent of information that a trainee has at a given time.

Finally, in the action stage, training designers must ensure that the training technology reacts to trainee input in a way that is predictable. This is doubly important.

Not only will proper design reinforce proper actions by the soldier trainee, but will also help to eliminate inappropriate actions from the trainee's behavioral repertoire.

Within the first (perceptual) stage of information processing, it is critical that soldiers attend to certain information, and recognize it as important. This task becomes more critical and more difficult as the number of competing information sources rises. In military environments, it is common for such a complex scenario to exist, because soldiers must be constantly aware of competing demands and information sources. As an example, it is frequently the case the there are multiple sources of communication information that may agree or conflict with each other during battle. Behavioral researchers such as Kahneman (1973) and Wickens (1984) have hypothesized about how human beings prioritize and act on competing information sources.

Kahneman, (1973) proposed that humans have a single pool of cognitive resources that they can allocate to one or more tasks. As more tasks draw from that pool, the available resources decrease, until no more resources are available. If one task alone is sufficiently demanding, there may be no way for a soldier to time-share other tasks, because all cognitive resources may be devoted to the first task. On the other hand, if there are several low-level tasks to perform, the trainee may be able to perform all of them effectively, as long as the total resource allocation does not exceed the limited capacity available. With practice, the extent of the resource pool may grow, so that more tasks may be accomplished concurrently. Assuming that this theory accurately depicts cognitive functioning, training designers need to create training scenarios that do not overload trainees, particularly at early stages of instruction. This is especially important when the consequences of task performance are high, as is the case with most military tasks.

Wickens and his colleagues (c.f., Wickens, 1984) proposed an alternative to the unitary resource model. They hypothesized that human beings have multiple pools of specialized cognitive resources available. For example, there could be a cognitive resource pool for verbal tasks, another for numeric tasks, and a third for spatial tasks. Based on this theory, trainees would be able to perform a number of tasks concurrently, as long as the tasks did not exceed the capacity of any particular pool of resources. Again, as trainees become more proficient at the tasks they learn, those tasks will tend to require fewer cognitive resources to complete (see Schneider & Schiffrin, 1977). This suggests that designers of training systems tailor the material to be learned with soldier capabilities. This idea has often been emphasized for the creation of training systems that adjust to learner's growing abilities, referred to as "adaptive training" systems (see Kelley, 1969).

<u>Approaches to Training</u>- Critical aspects of military training design include the setup, sequencing, and subsequent conduct of training. For soldiers to properly learn and retain information, it is important to design training sessions according to established research principles.

Researchers have concerned themselves with the proper spacing of training for many years. Bilodeau and Bilodeau (1961) reviewed much of the influential work in this area for motor skills. For years, researchers stressed that it was important to distribute training sessions over time, so that material may be learned and retained more readily.

However, the advantage of distributed training tends to disappear after the training period is over (see Epstein, 1949).

Interestingly, some learning literature suggests that, at least for the cognitive components of skill acquisition, optimal retrieval of memory may depend on the correspondence between practice spacing and the spacing between practice and test. In other words, it may be best to distribute practice sessions in a manner similar to the delay between the final practice session and the test session (see Anderson, 1995). Applying this guideline to military training exercises is difficult, because it is frequently difficult to estimate when training will be put to the test in battle. For that reason, training is often prescribed to accommodate monetary or temporal constraints instead of optimal learning.

In addition to proper practice distribution, tasks may need to be separated into component parts, so that practice sessions may concentrate on those components first, followed by increasingly complex combinations of parts, until the task may be practiced (and then performed) as a whole. Critical questions here concern how the task should be separated for training. McGeoch (1931) tackled this problem and concluded that the proper way to separate tasks for learning depended on a number of factors having to do with task performance context and experimental design. Since that time, numerous researchers have completed reviews of pertinent literature, but little consensus has emerged about how to separate tasks.

In general, it is fair to say that although task separation for training may yield a more learning-friendly environment, and may allow better learning of diverse subskills, identifying the best way to separate the overall task is still an elusive goal. Furthermore, there are significant issues concerning how to teach the reassembly stage (recombining task parts into performance of the whole task). AARs can be helpful because of the potential for illustrating how individual training strengths and weaknesses contribute to the overall success or failure of a global mission.

Training Issues - In addition to the cognitive issues discussed above, there are several training issues that must be addressed to optimize learning. These impact not only the current project, but are generally applicable to all efforts that involve training systems. One of the most important issues concerns how standards are set for successful training. Training researchers and designers have debated about the best ways to assess training effectiveness for years.

At a basic level, it is important to determine whether or not trainees have mastered the cognitive portions of the material effectively. For many, this is the most basic "acid test" of effective training. Trainees can demonstrate effective learning only if they understand the concepts that drive performance. Of course, there are a variety of types of learning that may be achieved. This fact was recognized early by Gagne (1985), and is illustrated in his taxonomy of learning capabilities that includes intellectual skills, verbal information, cognitive strategies, motor skills, and attitudes. Taxonomies are fluid tools, however, and may be modified for the particular task one is considering.

Although learning is certainly important, it is not a substitute for performance. In addition to understanding the intellectual aspects of the trained material, trainees must be able to apply what they have learned in an applied situation. This is particularly critical in the military, where many tasks are trained with the expectation that there will be subsequent, hands-on demonstration of learned skills.

Frequently, consideration of performance leads to the topic of transfer of training, where learned skills are expected to generalize positively to an operational environment. As noted by Patrick (1992), the goal of training is to maximize the amount of positive transfer from the training situation to the operational or "test" situation. However, positive transfer does not always take place. Trainees may learn strategies that they inappropriately apply in new situations, demonstrating "negative transfer" (see Luchins, 1942), or they may learn material that has no effect on the operational performance of the target task ("no transfer," see Proctor and Van Zandt, 1994). Various texts have included suggestions for promoting transfer of training, including specification of training objectives, increasing task variability during training and overtraining.

In the military, there is often a considerable time period that elapses between training periods. In fact, often soldiers may be trained to do a particular task, but may not have the opportunity to practice that task for months or even years after training. Therefore, one of the most significant challenges facing military training designers is to design training that results not only in immediate benefits to learning and performance, but that leads to long-lasting retention of learned skills. The long retention periods are another reason for the importance of AARs; specific and detailed performance feedback is important for trainees to retain pertinent information for long periods.

Naylor and Briggs (1961), and Hagman and Rose (1983) have provided influential reviews of literature pertaining to skill retention. The Hagman and Rose review is particularly important for the proposed project because of its emphasis on military tasks.

The available information concerning task retention suggests that three basic factors may lead to optimal task retention: 1) How well the trainees learn the task initially, 2) How long the retention interval is, and 3) Availability of practice periods within the retention interval.

As today's military has faced numerous cutbacks and financial strains, only the most cost-effective training methods are feasible. It is critical to consider that the term "cost effectiveness" refers not only to immediate cost savings, but to long-term savings associated with training technology reusability, and savings associated with multi-purpose hardware and software. It is a well-established fact that the use of simulation technology may help lower the overall costs of training. Cicchinelli, Harmon, Keller, and Kottenstette (1980), for example, determined that simulation offers a cost advantage when compared to the use of operational equipment for maintenance training.

Compared to physical simulators such as that evaluated by Cicchinelli et al. (1980), a benefit of immersive technology is its flexibility. Immersive training systems may be easily reconfigured to teach a variety of particular system-oriented skills. For example, a virtual environment to teach hostage rescue or supplies distribution skills may be easily reconfigured to support checkpoint operations skills. In contrast, physical mockups and simulators may require extensive cost and labor to meet varying skill needs.

Just as task performance is an important criterion for training success, it is critical that the skills trained in the military transfer positively to operational equipment. As pointed out by Holding (1965), trainers can ensure such transfer by optimizing the fidelity (physical and functional) of the training environment, and by properly scheduling training content and procedures. If training is properly designed and conducted, transfer to operational equipment will be guaranteed, and this will improve effectiveness in the field, and ultimately reduce costs.

Performance Feedback – Perhaps the most relevant variable for after action review is performance feedback. According to Kluger and DeNisi (1996), feedback is one of the most influential variables that affect training. In fact, Komaki, Heinzmann, and Lawson (1980) stress that training by itself is not enough to improve and maintain performance; rather, training plus feedback is the most effective strategy. Blum and Naylor (1968) point out that feedback can come from intrinsic sources within the individual learner, or extrinsic sources that are outside the learner. The reason for the effectiveness of performance feedback (knowledge of results) is generally attributed to the motivational and information-related benefits of such information.

Research regarding feedback effectiveness is also discussed in a later section specific to AARs; however, several research reviews have discussed the role of feedback in occupational training (Balzer, Doherty & Conner, 1989; Ilgen, Fisher & Taylor, 1979; Kluger & DeNisi, 1996). Generally, research findings suggest that feedback may be used effectively to influence learning. However, such a statement must be tempered by the detailed research findings concerning the administration of feedback. There are a number of variables that influence the effectiveness of feedback for learning.

Naturally, it is important for feedback content to be accurate and perceived as accurate by the learner. However, in addition to content, Ilgen, Fisher, and Taylor (1979) stress the importance of interpersonal dynamics between the feedback deliverer and recipient(s). Specifically, the expertise, reliability, credibility, power and attractiveness of the feedback deliverer and the recipients' perceptions of the feedback are important influences on learner acceptance and benefit. In their model, Ilgen et al. emphasized that learners often misperceive negative feedback, and that frequent feedback may not always be beneficial, as some learners may come to rely on it instead of developing their own performance evaluation skills. Ilgen et al.'s model seems to depict feedback as a social interaction process. For example, they suggest that individual needs of learners should be considered when selecting feedback, because competent individuals may benefit most from success-oriented information, whereas less competent individuals may need closer supervision and more specific feedback.

In a more recent meta-analytic review of the feedback literature, Kluger and DeNisi (1996) address some counterintuitive findings regarding feedback in the literature. For example, almost 40% of the studies they cited showed negative effects of feedback for task learning. Furthermore, some research has shown that the use of verbal praise alone may be worse than not using feedback at all. Kluger et al. suggest a number of interpersonal and content related reasons for these findings. Generally, they suggest that providers of performance feedback should strive to be specific. If feedback is too general, it may lead performers to adopt a superficial understanding of the task and may hinder transfer of training. In fact, they suggest that general positive feedback will be useful only in situations where effort is the strongest influence of performance and the learner suffers from low self-esteem. The best alternative is "cue feedback," in which learners are directed toward specific aspects of performance that should be improved. Kluger et al. note that if cue feedback is used, even negative feedback can be beneficial for learners who are committed to task mastery.

Clearly, the role of feedback is a complex one. However, if training designers consider the feedback issues cited above, in combination with the other issues related to trainee cognition and material content, the quality of the eventual training product will benefit.

Designing Virtual Environments to Facilitate Military Training

Considering the somewhat abbreviated list of training issues presented above, the proper design of training situations is no easy task. An interesting and helpful feature of virtual environment (VE) technology is that it is generally quite flexible. Although there is some debate about how best to train many skills, technology assists to optimize the training situation. As noted by Durlach and Mavor (1995), VEs offer qualities that are expressly called for in some of the most authoritative reviews on the use of simulation for training. VE designers have the potential to create a training environment that is optimized for human learning and information processing. In particular, virtual reality offers the following specific advantages for training curriculum design:

Standardization - An advantage of automated (or semi-automated) modes of instruction such as simulators and virtual reality systems is that the material taught is standard (or nearly so) for all trainees. While this is important for classroom instruction, it is doubly important for military environments, where the importance of accurate and reliable task execution cannot be understated.

Data Capture - As noted by Vreuls and Obermayer (1985), there are a host of issues associated with measuring performance in traditional task simulations. With the advent of immersive technology, behavioral scientists have the ability to collect a vast amount of performance data in raw and processed form. Such data may allow for real-time error correction and feedback, or for summative performance evaluation (after action review). One of the biggest challenges of that capability is determining how much data to collect, when to collect it, and how to use it to best optimize the learning process.

Some researchers (see Loftin & Savely, 1991) have proposed adding intelligent systems to the VR interface to allow for real-time and summative data collection, evaluation, and performance feedback. Such an approach is feasible, but depends largely on the existing theories of concerning the efficacy of feedback during training (Adams, 1987). The proposed research should help to address these needs by clarifying the role of a number of information presentation influences on feedback efficacy.

Fidelity - A major concern for training designers has been the realism, or fidelity represented by the virtual environment. Fink and Shriver (1978, p. 25) defined fidelity as the extent to which "trainees perceive training equipment as being a duplicate of the operational equipment and task situation." As noted by Hays and Singer (1989), fidelity may be characterized a number of ways.

Physical fidelity concerns the degree to which a training environment mirrors reality in a physical sense. That is, the physical aspects of the training environment (i.e., size, color, shape, etc. of objects) should resemble the "real-world" environment. This idea is largely based on Thorndike and Woodworth's (1901) "identical elements" theory of training transfer. That theory suggests that if physical fidelity is maximized, positive transfer of training will result, and learning will be guaranteed. This idea is inherent in Dale's (1969) "cone of experience," which classified instructional media according to the concreteness of the learning experiences as perceived by the learner.

In addition to physical fidelity, training environments show a variety of aspects related to functional fidelity. This type of fidelity is not concerned so much with physical properties of objects, but is more concerned with how they function. For trainees to learn properly, the objects in the training environment should behave in a manner similar to their counterpart objects in the actual (performance) situation (Allen, 1986). Again, any decrement here will result in a loss of training transfer.

A particularly critical problem for the design of training environments is the determination of desired fidelity. Designers of equipment simulators have debated this idea for years. Initially, many researchers believed that "more is better," and so made every attempt to maximize fidelity. Unfortunately, this increased the cost of training simulators, limiting their access to trainees (see Miller, 1954). Subsequently, debates ensued concerning the optimal amount of fidelity necessary for training devices. As pointed out by Patrick (1992), it is not an easily solved debate, because issues such as task criticality and the difference between physical and functional fidelity must be considered. As an example, Caro (1979), Koonce (1979), and Lintern (1987) debated the role of motion as a simulation characteristic for flight simulators for years, finally reaching a compromise point. In general, the determination of optimal training system fidelity is something that must be done specifically for each given training system.

<u>Control</u> - Research has shown that at times it is beneficial to vary the speed of a task for optimal training. Traditional simulation and virtual reality both offer the ability to manipulate aspects of the training situation so that important principles may be stressed during instruction. As an example, simulation and VR both allow training designers to simulate failures of equipment or processes, to train contingency reactions. Many researchers have lauded the idea of fault introduction (see White & Frederiksen, 1987, for example). Control of simulated features is likely of particular importance for AARs, because emphasis may be placed on environmental or task-related elements that are especially difficult or mission-critical.

<u>Presence</u> - Presence is a construct that describes the degree to which people are compelled that they are experiencing a separate environment (as opposed to being a passive recipient of a solely visual experience). Many researchers believe that a primary reason for the success of virtual reality for teaching is the opportunity for students to feel a part of a separate environment.

Virtual environments are not the only way that people experience presence. In the context of everyday activity, the phenomenon of presence occurs as a human senses information about the world, thereby achieving a sense of being present at a particular location, time, and space (Barfield, Zeltzer, Sheridan, & Slater, 1995). Barfield et al. speculate that presence is experienced as the cognitive centers in an individual's brain process visual, tactile, kinesthetic, olfactory, proprioceptive, and auditory sensory information to form an impression of the world.

Several researchers have discussed the benefits of virtual reality systems that are capable of producing presence. Held and Durlach (1992) propose that presence should enhance general performance in applications where individuals interact with synthetic, computer-generated worlds. Presence is also believed to benefit training situations where the development of motor functions is important (Held et al., 1992). Sheridan (1992) alleges that presence improves both sensorimotor and cognitive performance and increases the efficiency of training and planning. Although these predictions are intuitively attractive, they have not been addressed to a great degree by the research community. It is therefore necessary to empirically question whether presence has a positive or negative effect on learning that occurs in a virtual environment.

Given the current definitions of presence by researchers such as Sheridan (1992), Slater and Usoh (1993) and others, it follows that presence is an important attribute to maximize in VEs, particularly those used for training. As we evaluate the ability of soldiers to produce AARs following the review of recorded immersive training exercises, it is important to consider how realistic and convincing the virtual environment is. In this regard, it is likely that presence has strong ties to the notions of physical and functional fidelity, described earlier.

Team Training Using Virtual Environments

The definition of what constitutes a team has changed over the years. In the infancy of team research, Boguslaw and Porter (1962) defined the team relationship broadly, emphasizing the fact that the team involves interactions among members, machines and machine procedures. They also stressed the notion of a superordinate (system-related) goal.

In 1968, Klaus and Glaser made the distinction between small groups and teams. They stressed that team members have well-defined positions within the structure of the team, but members of small groups do not. Hall and Rizzo echoed these ideas in 1975, stating that teams are mission-oriented, possess formal structures and assigned roles, and require interaction among members.

Wagner, Hibbits, Rosenblatt, and Schulz (1977) further emphasized the structural component by stating that teams are "relatively rigid in structure and goal- or mission-oriented with the task of each team member well-defined (p. 14)". They also specified that a team could be composed of as few as two individuals who are associated together.

Several influential researchers have conceptualized teams as distinguishable sets of two or more individuals who interact interdependently and adaptively to achieve specified, shared, and valued objectives (Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986; Salas, Dickinson, Converse & Tannenbaum, 1992). Others have reinforced this definition (Briggs & Naylor, 1964; Dyer, 1986). From this, it is clear that the team definition has evolved into one that emphasizes interdependence and coordination.

Brannick, Prince, Prince, & Salas (1995) support this, defining team actions as "behaviors executed by two or more individuals as a function of coordinating requirements imposed by interdependent tasks in achieving common goals." Brannick et al. suggest that the essence of teamwork is team coordination (TC). However, little consensus has emerged regarding TC -- what it is, what particular skills it subsumes, and the proper measurement of those skills.

Much team research has attempted to isolate those factors that separate effective from ineffective teams. Klaus and Glaser (1968) suggested that cooperation and coordination are necessary components of team interaction. Subsequently, a distinction made by Morgan and his colleagues has enjoyed widespread support. They state that two sets of behaviors develop during team training: taskwork and teamwork (Morgan et al., 1986; Glickman, Zimmer, Montero, Guerette, Campbell, Morgan, & Salas, 1987). Taskwork includes behaviors that are necessary for the execution of specific tasks. In contrast, teamwork is more general across tasks, and includes interpersonal skills necessary to interact effectively with other team members.

Morgan et al. (1986) believe that teamwork behaviors fall into seven dimensions: giving suggestions or criticisms, cooperation, communication, team spirit or morale, adaptability, coordination, and acceptance of criticism or suggestions. Morgan et al.'s (1986) conceptualization is especially important because it appears to be the only example of prescriptive information for true "team" training before 1991 (Armstrong & Reigeluth, 1991).

Applying those ideas in a military setting, Oser, McCallum, Salas, and Morgan (1989) investigated what types of behaviors identified effective Navy teams. Helping other team members with tasks, making motivational statements, and reinforcing other team members were found to be characteristic of effective team members. Members of effective teams prompted teammates to recheck their work for mistakes, were thankful when teammates informed them of a mistake, gathered information effectively, and suggested ways to better detect errors. In contrast, members of ineffective teams often made negative comments about the team or the training procedure, and raised their voices while correcting other teammates.

Together, the results of these and other efforts have led many researchers to agree that three constructs are particularly important for team effectiveness: Communication, Cooperation, and Coordination (Brannick et al., 1995). More specifically, Tannenbaum, Smith-Jentsch, and Behson (1998) have proposed three means by which teams can promote and facilitate their own learning to continuously improve: preparing for a learning experience, managing its human resources, and learning from feedback regarding how well the team is performing.

Available team training research is relevant for the after action review process, because after action review reports are often constructed by teams of individuals who are responsible for separate but related aspects of soldier performance. It is therefore important that any examination of AAR feedback effectiveness include a component of team assessment.

After Action Reviews

Each of the issues discussed to this point has implications for the AAR process. AARs usually follow a fairly prescribed format. After a training simulation, trainers will commonly review the action that occurred during the exercise with the trainees. The goal is for trainees to improve their performance based on feedback that they receive from the AAR. As noted earlier, behavioral scientists have acknowledged the importance of feedback in training (Bilodeau, 1966). For trainees to glean maximal benefit from training exercises, it is imperative that they understand the aspects of their performance that that did or did not meet standards. Only after reviewing these things would a trainee be able to modify his or her performance to satisfy established criteria.

Past researchers have determined that performance feedback should have a number of qualities. It should follow task performance in a timely fashion, so that trainees may see the connection between their performance and the guiding criteria (Welford, 1968). Feedback should also be precise, to let trainees interpret it easily (Wheaton, Rose, Fingerman, Karotkin, & Holding, 1976). Feedback should be thorough, so that trainees may understand all aspects of their performance, and so that corrective comments can be understood within the context of the full training situation (Wheaton, et al., 1976). It should also be quantitative. Comments should focus on specific aspects of performance, rather than generalities (Magill & Wood, 1986).

In conventional military training exercises, soldiers often wage mock battles against an enemy. While doing so, they are observed by trainers who record positive and negative performance incidents. After the engagement, these incidents are presented to the soldiers, so that they might learn from their mistakes. This approach has proven beneficial; however, there are disadvantages. One problem is that because the observers who record performance are human, they are prone to bias. The parts of the battle that are noticed by one observer may differ from those noticed by another observer, because of past experiences, knowledge, and motivation. Therefore, the feedback that is provided (and the quality of it) will vary across observers.

A second limitation of current AARs is that observers can observe only a subset of activities that occur. This is especially true for modern warfare, because many of the tasks that are performed require significant information processing, communication, problem solving and decision making. Such intangibles may not be directly observable by human observers.

When researchers began to develop immersive simulations for military training, the focus was on technical capabilities and limitations. Systems for after action review were not a high priority as training designers focused mainly on getting the training hardware and software structured to function smoothly. However, as functionality improved, effort was devoted to the development of AAR capabilities.

By recording in real time the actions of trainees, the virtual AAR may serve as an objective action recorder, operating without bias or interruption. Consequently, there exists the capability to replay training exercises exactly as they happened so that trainers and participants may review performance from multiple points of view and overlay expert performance to observe discrepancies

To realize these benefits, trainers must be able to avoid the inherent limitations of using virtual environments for training. Being immersed in a virtual environment may be distracting. Instead of remaining focused on critical aspects of the task to be performed, trainees may focus on irrelevant aspects of the virtual environment or technical aspects (capabilities and limitations) of the simulation. In addition, as with other types of simulation, it is necessary for the trainees to disregard the idea that they are playing a game. Only then may they truly gain from the review process. Conduct of thorough AARs should discourage participants from adopting a "game" mindset.

The Dismounted Infantry Virtual After Action Review System

With these limitations in mind, researchers have devoted effort to create a suitable and effective after action review system for virtual environment training. The Dismounted Infantry Virtual After Action Review System (DIVAARS) developed by the Army Research Institute Simulator Systems Research Unit, in conjunction with the University of Central Florida's Institute for Simulation and Training, features many of the positive capabilities described earlier. The AAR system builds upon earlier simulation systems that have been equipment focused. It enables the review of performances by individuals, teams, and small-team leaders who operate on foot. This includes dismounted infantry, Special Operations Forces, members of law enforcement, and emergency first responders.

To date, DIVAARS has been subjected to limited usability testing. Specifically, Boone, Bliss, Headen, Lampton, Clark, and Martin (2003) required ROTC students to perform a number of simple tasks using a version of the system running on a laptop computer. Participants were required to consult the user's manual to accomplish the tasks. Subsequently, participants rated the general usability and accuracy of both the DIVAARS program and its manual. In general, the ratings of the user's manual were favorable, and participants were particularly supportive of the program itself. However, specific features within the DIVAARS graphical user interface were not evaluated, and the participants were not required to use the program to construct an actual after action review. The goal of the proposed research program is to continue this line of usability research. We plan to focus on the potential of DIVAARS to serve as a source of military mission replay information for soldiers tasked with preparing after action reviews. By doing so, we hope to contribute to the goals of the Virtual Environments for Training (VIRTE) program.

Research Plan

The specific plan for empirical investigation of feedback issues was predicated on current research goals envisioned by the Army Research Institute, and by NAVAIR, reflecting the mission of the VIRTE program. That mission, as stated within the VIRTE overview, included four goals: To enable effective training for sailors and marines in any context, to reduce training costs, to enable training for tasks that are impractical to train in conventional ways, and to improve the fidelity of simulation interfaces, improving their accessibility and realism.

To achieve these things, the proposed research constituted a teamed effort with personnel from NAVAIR-Orlando and the Army Research Institute (Orlando Field Unit) to investigate the role of training feedback as a part of after action reviews (AARs). The experimental team proposed a three-year plan to investigate three general influences on feedback effectiveness: timing, AAR content, and teaming. The specific issues to be addressed, and the methods by which they are addressed, were determined by mutual agreement with members of the NAVAIR and ARI research teams.

Year 1

During Year 1, we planned to investigate the role of recorded scenario replay timing on the effectiveness of after action review reports generated by soldiers. For the first task, we planned to conduct an experiment in which we manipulated the timing of the performance review. Using DIVAARS exercise scenarios generated at Fort Benning by ARI personnel, we planned to assemble active duty personnel at the Human Computer Interaction Laboratory (HCIL) on the ODU campus. We would then instruct them to review a number of scenarios (3), and to individually generate an after action review based on each scenario. In addition to generating the after action reviews, participants would complete a series of questionnaires to demonstrate the level of situation awareness, perceived workload, presence, and knowledge of scenario contents. We planned to use subject matter experts to judge AAR quality and comprehensiveness of the reviews generated by the participants.

In addition to timing, we anticipated that the duration of the scenario would be a powerful influence on soldiers' abilities to craft after action reviews. For that reason, we planned to have participants review a number of scenarios of varying duration but of comparable complexity. After doing so they would generate after action reviews in a manner similar to the first experiment. Again, the quality and comprehensiveness of the reviews would be measured, as well as situation awareness and perceived workload.

Year 2

During the second year of the project, we planned to examine the role of feedback presentation content. By manipulating the graphical display features within DIVAARS, we would present scenario information in a more or less detailed fashion. As an example, DIVAARS offers the capability to show or hide features such as entity tracks, entity identifier labels, and building features. We expected that manipulating such content features would reflect the role of cognitive resource reservoirs, so that predictions could be made about trainee performance. We planned to first manipulate macroenvironmental (global) interface features (those expected to influence a participant's perception of the entire battle space). Specifically, we expected features such as zoom level, graphics detail, use of color and motion, and point of view (e.g., entity point of view, "God's-eye" point of view, etc.) to be reasonable candidates. After action reports would be evaluated for their comprehensiveness and accuracy, and measures of situation awareness, presence, and perceived workload would also be collected and analyzed.

The second experiment in Year 2 was planned to complement the first. Continuing the theme of scenario content influence, experimenters planned to have participants generate after action reviews based on replayed DIVAARS scenarios with or without more detailed content features. Such features would be those that might alter participants' perception of individual scenario features (such as soldiers, buildings, or vehicles) or of battle subspaces (such as a city block, a forest region, or a group of buildings). Metrics similar to the other experiments would be collected and analyzed.

Year 3

In the last year of the project we planned to replicate the manipulations of timing and AAR presentation content completed in Years 1 and 2. However, the unit of analysis would be teams instead of individuals. Because of the logistical challenges associated with collecting data from teams of soldiers, only one major experiment was proposed for the third year. For that experiment, we planned to assemble squads of soldiers at the HCIL laboratory. After introducing them to the DIVAARS program and interface, we then were going to instruct them to review a series of DIVAARS scenarios, collectively generating written and/or oral after action review reports for each. The DIVAARS scenarios would differ in terms of content and timing feature availability, as in Years 1 and 2. As in prior experiments, data concerning situation awareness, perceived workload, and AAR comprehensiveness and accuracy were to be collected and analyzed.

Project Progress During the First Year

Literature Review – As a requirement of the contract, we devoted approximately six months (from contract start through July 1, 2005) toward assembling, organizing, and reading literature obtained for this project. The literature we acquired concerned the process of after action reviews, and how that process differs among the various military branches. The literature also included a considerable amount of theory and information about training methods for military tasks, and theory concerning human learning and feedback in general.

Card Sorting – We also completed a data collection and analysis effort that involved card sorting. Specifically, we created cards that presented adjectives describing an after action review system. Participants from ODU classes then sorted the adjectives into groups, so that we could obtain task analysis information for the activities involved in after action review. This activity lasted for approximately three months, from March through May of 2005.

Transcription and Review of Completed AARs From ARI – Within a Cooperative Research and Development Agreement (CRDA) we have in place with the PEO STRI field unit of the Army Research Institute, we assisted them by transcribing and reviewing a series of actual after action reviews conducted at military installations. Such data analysis was instrumental for us to understand the after action review process, and provided tremendous insight into the development of our research design for this project. Those activities lasted from April to approximately July of 2005.

Experimental Design Reformulation – Because of the difficulty encountered when attempting to recruit military participants, we were encouraged to develop a solution that would require fewer participants to achieve our experimental goals. To do this, we reformulated our target experimental design. This required approximately one month of work (during the June/July timeframe). The resulting experimental design is outlined below:

- Experiment 1a – Playback Timing within DIVAARS Interface
  - IV1 - Playback Speed (1:1; 3:1)
  - IV2 - Use of Playback Features (play only; full features)

- DV1 - AAR Preparation Scale (status, rationale, remedy) (SAGAT approach)
- DV2 - General and Specific Usability Questionnaire
- DV3 - Situation Awareness Questionnaire (SART)
- DV4 - Workload (SWORD)
- DV5 - Subject Matter Expert's Rating of AAR Appropriateness, Comprehensiveness, Accuracy

- Experiment 1b - Content Features within DIVAARS Interface
  - IV1 - Access to Tables (yes, no)
  - IV2 - Views (entity view, battle commander view, top-down entity view)
  - DVs - Same as in Experiment 1a

- Piggybacking Approach (Only 20 participants needed):
  - 1a - All capabilities w/ Limited playback vs. Limited Playbacking Speed
  - 1b - All vs. No tables (regular playback; all features)
  - 2a - All vs. Entity view only
- Project 1a time scale:

15 min	15 min	15 min	15 min	30 min	30 min	30 min	30 min
1:1 Play	3:1 Play	1:1 All	3:1 All	1:1 Play	3:1 Play	1:1 All	3:1 All

Note: during each of the breaks, a questionnaire is to be administered lasting 15-30 min. Total time: 7 hours + 1 hour break = 8 hours

- Logistical Alternatives:
  - Local approach: facilitator, confederates co-located
  - Teletechnet approach: facilitator at NPS; confederate audience at ODU. All activities at NPS (videotaped and sent to ODU)

Participant Recruitment Timeline – Because much of our efforts during the first year of this project centered around acquisition of military participants, it is important to document the activities that occurred. Below is a chronological listing of those activities:

April 12, 2005: After discussing an approach with personnel from ONR, NAVAIR/Orlando, and VMASC, we approached Major Rodney Choi (Expeditionary Warfare School at Quantico) using assistance from Dr. Nathan Bailey at VMASC. MAJ Choi stated that he didn't think Quantico marines would have the time to participate. Maj. Choi mentioned that they were experiencing a high turnover of personnel during summer 2005 and that he did not think they would have any significant number of personnel available for the time needed, particularly as we desired to do follow-on work in the Fall of 2005. They were already involved in a number of projects in which we have been directed to participate. They had no more "white space" on the calendar for a voluntary project.

May 5, 2005: We next contacted Major Sutton, at Quantico's MOUT training facility (through Dr. Nathan Bailey and Bob Armstrong at VMASC). MAJ Sutton indicated that, though he was interested, there was no room on the docket to support this project.

May 30, 2005: We then contacted Dr. Rudy Darken at the Naval Postgraduate School (NPS) in Monterrey, CA, through LTC Joseph Cohn of ONR. Personnel at that school (William Becker) expressed willingness to help, provided we could organize the project and be ready to begin data collection activities by the start of the Fall semester. Three problems put this option at risk: The geographical distance of participants, the unavailability of DIVAARS software and associated scenarios from ARI, and technical challenges of AAR administration over TELETECHNET. However, we kept NPS as a backup possibility.

June 15, 2005: We contacted personnel at Quantico again, this time through K.C. Pfluger (upon recommendation from ONR). After about a month, K.C. indicated that Quantico was reluctant to sign on due to the numbers needed and time requirement. He suggested it would probably be best to go with NPS.

July 5, 2005: After revising our experimental design (see above), we reapproached K.C. Pfluger. Our experimental team had found a way to reduce the numbers needed, and asked K.C. to reapproach Quantico with those numbers. He did, but indicated that they were still not receptive. At this point, we decided to attempt more local sources, with NPS as a backup plan.

August 5, 2005: We contacted personnel at the Joint Forces Staff College (specifically, COL Fred Kienle and his assistant, LtCol. Terri Wilcox) about the possibility of testing students or faculty members from JFSC. LtCol. Wilcox mentioned that there were several classes, totaling 250 every 10 weeks. However, after checking with the individual instructors, she stated that no students were available from CPT Boatright's class or CPT McCloney's class, and that COL Kienle's students were not available either. She did mention the possibility of using JFSC faculty as participants. However, the number of faculty members available would be limited (n=approx. 5). As a backup, Wilcox suggested contacting personnel at Fort Eustis (specifically, a retired contractor named Dave Stevenson at 443-5802, or LtCol. Scott Waterman). She also suggested approaching either J5 or J3.

August 25: As an entrée into J5, J3, J7, or J9, we contacted Col. Michael LoQuasto and Col. Jerry Wood (Army War College Fellows at the Virginia Modeling, Analysis, and Simulation Center). They instead brought up the possibility of recruiting Army National Guard Units. They mentioned that the 29th Division Infantry in the Hampton Roads area may be a possibility. They also mentioned "FA" units in the area. However, after describing our needs, they suggested the time commitment (6-8 hours) may be excessive for National Guard units who are active only one weekend per month. They did, however, promise to circulate the word. One question they had was whether volunteers could get anything in return, such as access to the DIVAARS software. As we were not in a position to offer this, we decided to continue checking other options.

August 28: Finally, we contacted Ret. Col. Paul Nichols (through Dr. Sherrie Jones at NAVAIR Orlando), who was present as a marine Subject Matter Expert at the Arlington team training workshop. He indicated willingness to meet and work to access local marines. As of September 9th, however, he was unable to meet with us. We continued to attempt a meeting. However, shortly after this we were informed by telephone that the project was not going to be funded after the first year.

Discussion and Conclusions

We anticipated that our analyses would allow the precise determination of the influence of each of the variable sets indicated: timing of replay features, scenario content review features available within the DIVAARS interface, and the use of teams to perform after action reviews.

Examination of each of these variable sets is a crucial step for the realization of basic and applied goals. As is evident from the research literature, there is little clear consensus regarding the influence of feedback as it is utilized within after action review systems. Our results could assist personnel from NAVAIR-Orlando to fill in gaps within their Training Intervention Matrix (TIMx) so that a comprehensive picture of feedback influences and effects may be available. More broadly, our findings should be publishable within refereed training journals to assist the scientific community to understand the role of feedback and to prepare explanatory and predictive theories regarding it.

Our research should also assist military training experts to understand the influence of certain variables toward AAR effectiveness. Such information could ultimately result in contributions toward overall training effectiveness. More specifically, it could demonstrate the utility of the DIVAARS program and highlight any usability obstacles that may need to be anticipated and corrected before full implementation of the program.

Unfortunately, given the retraction of funding, we were not able to complete much of our planned effort. However, we remain eager to complete the research should funding be made available at some future point.

References

Adams, J.A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101(1), 41-74.

Allen, J. (1986). Maintenance training simulator fidelity and individual differences in transfer of training. Human Factors, 28 (5), 497-509.

Anderson, J.R. (1995). Learning and memory: An integrated approach. New York: John Wiley & Sons.

Armstrong, R.B., & Reigeluth, C.M. (1991). The TIP theory: Prescriptions for designing instruction for teams. Performance Improvement Quarterly, 4(3), 13-40.

Barfield, W., Zeltzer, D., Sheridan, T., & Slater, M. (1995). Presence and performance within virtual environments. In W. Barfield, & T. A. Furness III (Eds.), Virtual environments and advanced interface design. New York: Oxford University.

Bilodeau, E. A. (1966). Supplementary feedback and instructions. In E. A. Bilodeau (Ed.), Acquisition of Skill. New York, NY: Academic.

Bilodeau, E.A., & Bilodeau, I.M. (1961). Motor skills learning. Annual Review of Psychology, 12, 243-280.

Blum, M. L., & Naylor, J. C. (1968). Industrial psychology: Theoretical and social foundations. New York: Harper & Row.

Boguslaw, R., & Porter, E.H. (1962). Team functions and training. In R.M. Gagne (Ed.), Psychological principles in system development. New York: Holt, Rinehart, and Winston.

Boone, C.M., Bliss, J.P., Headen, E.N., Lampton, D.R., Clark, B., & Martin, G. (2003). Usability evaluation of a virtual after action review system. Proceedings of the Hunstville Simulation Conference. October 29-31: Huntsville, AL.

Briggs, G.E., & Naylor, J.C. (1964). Experiments on team training in a CIC-type task environment. (Technical Report NAVTRADEVCEN 1327-1). Port Washington, NY: US Naval Training Device Center.

Caro, P.W. (1979). The relationship between flight simulator motion and training requirements. Human Factors, 21, 493-501.

Cicchinelli, L.F., Harmon, K.R., Keller, R.A., & Kottenstette, J.P. (1980). Relative cost and training effectiveness of the 6883 three-dimensional simulator and actual equipment. (Technical Report No. 80-24). Brooks Air Force Base, TX: Air Force Human Research Laboratory.

Dale, E.A. (1969). Audiovisual methods in teaching. (3rd ed.). New York: Holt, Rinehart, and Winston.

Durlach, N.I., & Mavor, A.S. (eds)(1995). Virtual reality: Scientific and technological challenges. Washington, D.C.: National Academy Press.

Dyer, J.L. (1986). Annotated bibliography and state-of-the-art review of the field of team training as it relates to military teams. (Research Note 86-18). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences

Endsley, M. (1995). Towards a theory of situation awareness. Human Factors, 37(1), 32-64.

Epstein, B. (1949). Immediate and retention effects of interpolated rest periods on learning performance. New York: Teaching College Continuing Education.

Fink, C.D., & Shriver, E.L. (1978). Simulators for maintenance training: Some issues, problems, and areas for future research. (Technical Report 78-27). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Gagne, R.M. (1985). The conditions of learning and theory of instruction. New York: CBS College Publishing.

Glickman, A.S., Zimmer, S., Montero, R.C., Guerette, P.J., Campbell, W.J., Morgan, B.B., & Salas, E. (1987). The evolution of teamwork skills: An empirical assessment with implications for training. (Technical Report No. NTSC 87-016). Arlington, VA: Office of Naval Research.

Hagman, J.D., & Rose, A.M. (1983). Retention of military tasks: A review. Human Factors, 25(2), 199-213.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). Amsterdam: North-Holland.

Hays, R.T., & Singer, M.J. (1989). Simulator fidelity in training system design. New York: Springer-Verlag.

Held, R. M., & Durlach, N. I. (1992). Telepresence. Presence: Teleoperators and Virtual Environments, 1, 109-112.

Holding, D.H. (1965). Principles of training. Oxford: Pergamon.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349-371.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kelley, C.R. (1969). What is adaptive training? Human Factors, 11, 547-556.

Klaus, D.J., & Glaser, R. (1968). Increasing team proficiency through training: 8. Final Summary Report AIR E 1-6/68FR. Washington, D.C.: American Institutes for Research.

Kluger, A. N. & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. Current Directions in Psychological Science, 7, 67-72.

Komaki, J., Heinzmann, A.T., & Lawson, L. (1980). Effect of training and feedback: Component analysis of a behavioral safety program. Journal of Applied Psychology, 65, 261-270.

Koonce, J.M. (1979). Predictive validity of flight simulators as a function of simulator motion. Human Factors, 21, 215-223.

Lintern, G. (1987). Flight simulation motion systems revisited. Bulletin of the Human Factors Society, 30(12), 1-3.

Loftin, R.B., & Savely, R.T. (1991). Applications of Intelligent Computer-Aided Training. Proceedings of the Conference on Computers in Aerospace VIII. Baltimore, MD, October 22-23.

Luchins, A.S. (1942). Mechanization in problem solving: The effect of Einstellung. Psychological Monographs, 54, 248.

Magill, R.A., & Wood, C. A. (1986). Knowledge of results precision as a learning variable in motor skill acquisition. Research Quarterly for Exercise and Sport, 57, 170-173.

McGeoch, G.O. (1931). Whole-part problem. Psychological Bulletin, 28(10), 713-739.

Miller, R.B. (1954). Psychological considerations in the design of training equipment. (Final Report 54-563). Wright-Patterson Air Force Base, OH: Wright Air Development Center.

Morgan, B.B., Glickman, A.S., Woodard, E.A., Blaiwes, A.S., & Salas, E. (1986). Measurement of team behaviors in a Navy environment. (Tech Report NTSC TR-86-014). Orlando, FL: Naval Training Systems Center.

Naylor, J.C., & Briggs, G.E. (1961). Long-term retention of learned skills: A review of the literature. (Technical Report No. 61-390). Wright-Patterson Air Force Base, OH: Behavioral Sciences Laboratory.

Oser, R., McCallum, G.A., Salas, E., & Morgan, B.B. (1989). Toward a definition of teamwork: An analysis of critical team behaviors. (Technical Report No. TR-89-004). Orlando FL: Naval Training Systems Center.

Patrick, J. (1992). Training: Research and practice. London: Academic Press.

Proctor, R.W., & Van Zandt, T. (1994). Human factors in simple and complex systems. Boston: Allyn & Bacon.

Rasmussen, J. (1986). Information processing and human-machine interaction: An approach to cognitive engineering. Amsterdam: North-Holland.

Salas, E., Dickinson, T.L., Converse, S., & Tannenbaum, S.I. (1992). Toward an understanding of team performance and training. In R.W. Swezey and E. Salas (Eds.), Teams: Their training and performance (pp. 3-29). Norwood, NJ: Ablex.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 1-66.

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments, 1, 120-125.

Slater, M., & Usoh, M. (1993). Presence in immersive virtual environments. Proceedings of the IEEE Virtual Reality International Symposium (VRAIS).

Tannenbaum, S. I., Smith-Jentsch, K. A., & Behson, S. J. (1998). Training team leaders to facilitate team learning and performance. In J. A. Cannon-Bowers & E. Salas (Eds.), Decision Making Under Stress: Implications for Individual and Team Training (pp. 247-270). Washington D.C.: American Psychological Association Press.

Thorndike, E.L., & Woodworth, R.S. (1901). I. The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. III. Function involving attention, observation, and discrimination. Psychological Review, 8, 247-261, 384-395, 553-564.

Vreuls, D., & Obermayer, R.W. (1985). Human-system performance measurement in training simulators. Human Factors, 27(3), 241-250.

Wagner, H., Hibbits, N., Rosenblatt, R.D., & Schulz, J.R. (1977). Team training and evaluation strategies: State of the art. (Technical Report No. TR-771-1). Alexandria, VA: Human Resources Research Organization.

Wheaton, G. R. Rose, A. M., Fingerman, P. W., Korotkin, A. L., & Holding, D. H., (1976). Evaluation of the effectiveness of training devices: Literature review and preliminary model. (Research Memorandum No. 76-6). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Welford, A. T. (1968). Fundamentals of skill. London: Methuen.

White, B.Y., & Frederiksen, J.R. (1987). Qualitative models and intelligent learning environments. In Lawler, R.W., and Yazdani, M. (eds.), Artificial intelligence and education, Volume 1: Learning environments and tutoring systems. Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (eds.), Varieties of Attention (pp. 63-102). New York: Academic Press.

Witmer, B. G., & Singer, M. J. (1994). Measuring immersion in virtual environments. (Technical Report 1014). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.